DOCUMENT RESUME

ED 190 641 TM 800 436

AUTHOR Plake, Barbara S.; And Others
TITLE Number Right and Elimination Score as a Function of Item Arrangement, Knowledge of Arrangement, and Test Anxiety.
PUB DATE Apr 80
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 8-10, 1980). The appendix was removed because it contains copyrighted test instruments.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Difficulty Level; Higher Education; Multiple Choice Tests; Quantitative Tests; *Scores; Scoring Formulas; *Test Anxiety; *Test Format; Test Items; *Test Wiseness
IDENTIFIERS *Elimination Scoring; *Number Right Scoring; Test Instructions

ABSTRACT

Number right and elimination scores were analyzed on a 48-item college level mathematics test that was assembled from pretest data in three forms by varying the item orderings: easy-hard, uniform, or random. Half of the forms contained information explaining the item arrangement and suggesting strategies for taking the test. Several anxiety measures (state, trait, facilitating, and debilitating) were administered, along with the experimental forms of the mathematics test, to 97 undergraduates. Multivariate analysis of variance revealed no significant effects of the independent variables (order, knowledge, and anxiety condition) on either cognitive dependent measure (number right or elimination scores). Multivariate analysis of covariance, using measures of anxiety as multiple covariates, also failed to show any significant influence on the cognitive dependent measures or a set of perceptional dependent measures. Results tentatively suggest that neither test scores nor attitudes were influenced by item order, knowledge of item arrangement, or anxiety level. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190641

Number Right and Elimination Score as a Function of Item Arrangement, Knowledge of Arrangement and Test Anxiety

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Barbara S. Plake

Patricia A. Thompson

Stephen Lowry

University of Nebraska--Lincoln

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B. Plake

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800436

Paper presented at the National Council on Measurement in Education Annual Meeting, Boston, 1980

Number Right and Elimination Scores as a Function of Item Arrangement, Knowledge of Arrangement and Test Anxiety

Barbara S. Plake

Patricia A. Thompson

Stephen Lowry

University of Nebraska--Lincoln

ABSTRACT

Number Right and Elimination scores were analyzed on 48-item college level mathematics exam that was assembled from pretest data in three forms by varying the item orderings: easy-hard, uniform, random. Half of the forms contained information explaining the item arrangement and suggesting strategies for taking the test. Several anxiety measures were administered in a balanced order along with the experimental forms to 97 undergraduates. Multivariate ANOVA revealed no significant effects of the independent variables (order, knowledge, anxiety condition) on either cognitive dependent measure (NR or EL) ($\alpha = .05$). Multivariate ANCOVA, using measures of anxiety as multiple covariates, also failed to show any significant influence ($\alpha = .05$) on the cognitive dependent measures or a set of perceptional dependent measures. Results suggest that neither test scores nor attitudes are influenced by item order, knowledge thereof, or anxiety level, although factors contributing to the non-significance may have been inherent in the research design because of inadequate power and/or student performance motivation.

Number Right and Elimination Scores as a Function of Item Arrangement, Knowledge of Arrangement and Test Anxiety

Barbara S. Plake

Patricia A. Thompson

Stephen Lowry

University of Nebraska--Lincoln

Multiple-choice achievement test results are used to estimate student's achievement throughout the educational process. Potentially confounding factors to the validity of these scores include item arrangement, response set, test anxiety, and scoring method. Research studies have typically used some subset of the four item arrangements (easy-hard, hard-easy, uniform, and random). The results generally support easy-hard ordering for speeded tests but have shown no practical differences for power tests (Flaugher, 1968; Brenner, 1964). Studies combining test anxiety and item arrangement show conflicting results (Hambleton & Traub, 1974; Marso, 1970). Response set, or expectation by the students of the order of the items, was not shown to effect test performance (Plake & Seefeldt, 1978), although researchers have identified expectation of item arrangement to be a factor in test performance (Cronbach, 1964; Hambleton & Traub, 1974). The most common scoring method for multiple-choice tests is number right (NR). This scoring method does not give credit for partial or incomplete knowledge. Another scoring system, elimination scoring (Coombs, Milholland and Womer, 1956), has been designed to be more sensitive to student achievement, however, research has not confirmed (Hakstain & Kansup, 1975). Also, since the procedure is

complex and unique for the test taker, performance on a test score by the elimination method may be depressed for a student who has a high test taking anxiety.

Instruments

A 48 item power mathematics test was assembled from items of the ACT: College Mathematics Placement Program to represent a range of items to cover content from elementary mathematics through college algebra. The item difficulty indices ranged from .15 to .84 with a median value of 0.55. The item discrimination indices (point biserial) ranged from .15 to .82 with a median = .53.

Three test forms were developed from the 48 item multiple choice test. In the first form (E-H), the items were simply arranged in ascending order based on their item difficulty indices. Items in the second form were ordered in a uniform fashion, with cycle equal to 4 (Kleinke, 1979). The uniform (or spiral cyclic) order contained 12 sets of 4 items; within each 4 item set, the items were arranged in ascending order of difficulty. The first item was from the upper quarter of the distribution of items based on their item difficulty, the second one was from the next quarter, and so on. The 4th item in each set was selected from the 12 most difficult items in the 48 item pool. The 12 sets of items were themselves placed in ascending order based on the combined item difficulties of the items within each set. The first cycle consisted of the easiest item from each of the 4 quarters of the item difficulty distribution (i.e., items that ranked 1st, 13th, 25th, & 37th). The second cycle was formed by selecting the next easiest items within each quarter of the item distribution, and therefore was formed by the 2nd, 14th, 26th and 38th items in terms of item difficulty ranking. The 12th cycle, therefore, consisted of the items

with rank order on item difficulty of 12th, 24th, 36th and 48th The third test form (random) was established by placing the 48 multiple choice items in a random arrangement. In the directions which were developed to accompany each test booklet, half contained additional information explaining the order of the items in the form. In the other half of the test booklets, the directions addressed the scoring and marking methods only but omitted information about the arrangement of the items in the form. Therefore, six experimental forms were developed: 3 item arrangements x 2 conditions of knowledge of arrangement.

Three measures of anxiety were used in this study: Alpert and Haber's 19 item Achievement Test Anxiety Test (AAT), Spielberger's STATE and TRAIT Anxiety Inventories (STAI-STATE, STAI-TRAIT). AAT is reported to have two subscales, Facilitating and Debilitating (Alpert & Haber, 1960). Copies of the instruments used in the study can be found in the appendix.

Procedure

The six experimental forms were administered to 97 undergraduates enrolled in an educational psychology class at the University of Nebraska-Lincoln, Spring, 1979. The students volunteered to take part in the study and received some course credit for participation in the research study. The students were informed of the basic intent of the study and that they were free to withdraw from the study, without prejudice, at any time. All of the subjects signed their informed consent form and none elected to withdraw from the study. The test materials were administered to all the students in one 3 hour evening session. The instruments were assembled into packets and distributed in a random fashion to the students as they arrived for the research project. Each packet contained the three

anxiety measures (AAT, STAI-TRAIT, STAI-STATE), arranged in balanced order, followed by the experimental form of the mathematics test. Directions on the test form instructed the students to 1) select the best alternative for each item, 2) eliminate the alternatives that were incorrect for each and, 3) give an estimate for each item of their confidence of their selection by using a 1-5 scale ("I guessed" to "I am sure of my answer"). Additional questions were added to the end of each test form to assess students' perceptions of the test's difficulty and their expected performance. Immediately after the experimental form, all students were readministered the STAI-STATE anxiety test.

Analysis

Anxiety condition was defined by using the Facilitating and Debilitating scores from the AAT in a manner similar to that used by Towle and Merrill (1975). First, the difference score (D) for each subject was determined by subtracting his/her Debilitating score from Facilitating score. The 97 D values were then ranked in descending order of magnitude. The students in the upper quarter were identified as being in the high facilitating--low debilitating (HL) anxiety group. Students in the lower quarter became members of the low facilitating/high debilitating (LH) anxiety group. For the remaining half of the students, whose D scores were neither in the upper or lower quarter, the sum of the Facilitating and Debilitating subscales (S) was determined. These S values were then ranked and students in the upper 1/2 of this group became members of the high facilitating/high debilitating group (HH). The remaining students were classified as low facilitating/low debilitating in anxiety condition (LL).

The first analysis was a multivariate ANOVA (3 x 2 x 4). In this

analysis, number right (NR) and elimination (EL) scores were the dependent measures; item arrangement (3 levels), knowledge of arrangement (2 levels), and test anxiety condition from AAT (4 levels) were the independent measures. A second analysis was used to examine student perception scores (confidence, perceived difficulty, expected performance) by using item arrangement and knowledge of arrangement as the independent variables, considering STAI-TRAIT a covariate. A 3-factor repeated measures ANOVA was then used to investigate if any of the experimental forms significantly altered scores from two administrations of the STATE Scale of the STAI. Multivariate multiple regression of Debilitating, Facilitating, STATE, and TRAIT anxiety measures on NR and EL scores was used to further investigate the relations between anxiety and test performance. Finally, reliability indices (KR_{20}) for the six experimental forms were compared.

Results

The means and standard deviations for the NR, EL, confidence, perception of difficulty, and perception of performance scores for combinations of orderings and knowledge of orderings are shown in Table 1.

- - - - - - - - - - - - - - - - - - - -
Insert Table 1 about here
- - - - - - - - - - - - - - - - - - - -

None of the multivariate tests for the main effects of anxiety condition, knowledge of ordering, or item arrangement, or any of one higher order interactions was significant ($\alpha = .05$). Results can be found in Table 2. When the dependent measures were considered in a stepdown fashion, anxiety condition by knowledge of ordering did produce a significant effect ($\alpha = .05$) on NR scores. This interaction may not be meaningful, however, since the multivariate test for anxiety condition by knowledge of ordering was not

significant. An inspection of the cell means for anxiety condition and knowledge of ordering showed that the performance for the HH anxiety condition scores were depressed under the informed condition for knowledge of ordering, ($\bar{X}_{HL}$ = 22.13, $\bar{X}_{LH}$ = 28.27, $\bar{X}_{HH}$ = 18.18, $\bar{X}_{LL}$ = 27.45). For subjects in the noninformed condition of knowledge of item ordering, the performance was depressed for those subjects in the LL anxiety condition ($\bar{X}_{HL}$ = 25.45, $\bar{X}_{LH}$ = 27.42, $\bar{X}_{HH}$ = 25.85, $\bar{X}_{LL}$ = 19.54). Also for this data, the highest NR scores were obtained by students in the LH anxiety condition for both informed and non-informed ordering conditions.

- - - - - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - - - - -

The correlations between test performance measures (NR and EL) and the perceptual measures of confidence, difficulty, and performance are shown in Table 3. All of these correlations are significant (α = .05), indicating test performance was directly related to student perceptions. A two factor MANCOVA was undertaken to analyze the effects item order and knowledge of order on each of the perception scores when TRAIT anxiety was used as a covariate. The results are summarized in Table 4 and demonstrate no significant effects for item order and knowledge of order, or item by order interaction when TRAIT anxiety is used as a covariate on confidence, perceived performance, or perceived difficulty.

- -

Insert Tables 3 & 4 about here

- -

Since anxiety condition did not significantly effect test performance,

an additional analysis was run to further investigate the effects on test performance (NR & EL) of item order and knowledge of orderings when multiple anxiety measures are used as covariates. A multivariate analysis of covariance was used, with AAT Facilitating and Debilitating scores, and Spielberger's STATE and TRAIT measures as covariates on both NR and EL scores. The set of anxiety variables accounted for significant percents of variance on NR and EL scores ($\alpha = .05$), but the percents of variance accounted for by the anxiety measures were not impressive (11% and 10% for NR and EL, respectively). No significant effects for item order, knowledge of order, or item by knowledge interaction were found. The MANCOVA results are summarized in Table 5.

- - - - - - - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - - - - - - -

The effect of item arrangement and knowledge of arrangement was further investigated using a repeated measures ANOVA on pre/post STATE anxiety scores. Although, STATE anxiety scores were significantly lower ($\alpha = .05$) after the testing experience, this reduction did not appear to be a function of experimental form, even though time by item ordering interaction approached significance ($p < .10$). A summary of this analysis is presented in Table 6.

- - - - - - - - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - - - - - - - -

KR_{20} reliabilities were calculated for each of the six experimental forms. These values are displayed in Table 7. The internal consistency

values were tested for population equality using a procedure described by Feldt (1969). Random ordering for both informed and non-informed condition produced the lowest KR_{20} values, although .88 and .89 reliabilities are satisfactory values. These reliability indices are significantly lower than either of the other reliabilities for the non-informed tests (Forms 4 and 5) and the Uniform ordering under informed condition (Form 2). All of the reliabilities, however, are within acceptable bounds for a classroom achievement test.

- - - - - - - - - - - - - - - - - - - -

Insert Table 7 about here

- - - - - - - - - - - - - - - - - - - -

The results suggest that performance on power multiple-choice items, whether scored simply as number right or by an elimination procedure, are not effected by item ordering or knowledge thereof, even when student anxiety level is taken into account. Test developers may therefore find their energies better spent in item development without overdue concern for the item ordering for a power test. These conclusions generalize to elimination scoring procedures and students of varying levels of test anxiety. However,the research nature of the study may have not provided sufficient motivation for performance on the mathematics test. These results, therefore, may not generalize to the actual classroom or standardized test situation where student's performance may be more important to the student. Furthermore, the research design and sample size may not have provided adequate power for the analyses selected to investigate the effects of item arrangement, knowledge of arrangement and anxiety condition on NR and EL scores. Trends in the data do support the presence of some

possible effects. Therefore, conclusions based on the results of this study should be tempered by the knowledge that the lack of significant effects may have been due in part to insufficient power and/or motivation in the research design.

References

Alpert, R., Haber, R. N. Anxiety in academic achievement situation, Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Brenner, M. H. Test Difficulty, reliability, and discrimination as functions of item difficulty order. Journal of Applied Psychology, 1964, 48 (2), 98-100.

Coombs, C. H., Milholland, J. E., & Womer, F. B. The assessment of partial knowledge. Educational and Psychological Measurement, 1956, 16, 13-37.

Cronbach, L. J. Response set and test validity, Educational and Psychological Measurement, 1964, 6, 475-494.

Feldt, L. S. A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests, Psychometrika, 34: 363-373, 1969.

Flaugher, R. L., Melton, R. S., & Meyer, C. T. Item rearrangement under typical test conditions, Educational and Psychological Measurement, 1968, 28, 813-824.

Hakstain, A. R., & Kansup, W. A comparison of several methods of assessing partial knowledge in multiple choice tests II, Testing Procedures, Journal of Educational Measurement, 1975, 4 (12), 231-240.

Hambleton, R. K., & Traub, R. E. The effects of item order on test performance and stress, Journal of Experimental Education, 1974, 43 (1), 40-46.

Kleinke, D. J. "Item Order, Response Format, and Examinee Sex and Handedness and Performance on a Multiple-Choice Test," Paper presented at the annual convention of the National Council on Measurement in Education, San Francisco, 1979.

Marso, R. N. Test item arrangement, testing time, and performance. Journal of Educational Measurement, 1970, 7 (2), 113-118.

Plake, B. S., & Seefeldt, F. M. The effect of item arrangement and Response Set on Total Test Score and on Students' Opinion of the test, paper presented at the annual meeting of Rocky Mountain Educational Research Association, Albuquerque, N.M., 1978.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. Manual for the State-Trait Anxiety Inventory, Palo Alto: Consulting Psychologist Press, 1970.

Towle, N. J., Merrill, P. F. Effects of anxiety type and item-difficulty sequencing on mathematics test performance, Journal of Educational Measurement, 1975, 4 (12), 241-250.

Table 1

Means and Standard Deviations

Number Right (NR), Elimination (EL) scores, Confidence (CONF),

Perceived difficulty (PD), Perceived performance (PP) scores

			Order					
			E-H		Uniform		Random	
			$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
		NR	25.5	9.7	23.3	10.4	22.6	9.3
		EL	55.9	46.5	54.9	43.7	59.9	43.9
	Yes	CONF	112.7	30.5	119.0	30.9	137.5	46.1
		PD	3.2	1.2	2.5	0.5	2.3	1.1
		PP	2.5	0.8	2.0	0.9	2.5	1.2
Knowledge			$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
		NR	23.0	10.5	23.3	10.4	27.1	9.0
		EL	56.6	47.1	55.9	48.8	65.2	42.6
	No	CONF	122.8	38.2	106.0	36.0	122.6	21.2
		PD	2.8	1.2	3.3	1.4	2.7	1.1
		PP	2.7	1.0	3.2	0.7	2.9	0.7

Table 2

MANOVA Summary Table

Number Right (NR), Elimination (EL)

Source	Multivariate		Stepdown F	
	F	P	NR	EL Eliminating NR
Order	$F_{4,144} = .25$	.91	.21	.29
Knowledge	$F_{2,144} = .05$	.95	.10	.003
Anxiety Condition	$F_{6,144} = 1.28$	.27	1.60	.99
Order x Knowledge	$F_{4,144} = 1.75$	.14	1.63	1.88
Order x Anx Cond	$F_{12,144} = 1.06$	.40	1.16	.97
Knowledge x Anx Cond	$F_{6,144} = 1.64$	.14	2.79*	.58
Order x Knowledge x Anx Cond	$F_{10,144} = .54$	.86	.41	.67

* $p < .05$

Table 3

Correlations between Test Performance (NR and EL)

and Perceptual Scores (CONF, PD, PP)

	NR	EL	CONF	PD	PP
NR	1.0				
EL	.93	1.0			
CONF	.75	.78	1.0		
PD	.72	.67	.75	1.0	
PP	.56	.56	.53	.54	1.0

Table 4

MANCOVA Summary

Confidence, Perceived Difficulty, Perceived Performance

with TRAIT as covariate

Source	Multivariate F	P	Stepdown F: Confidence	Stepdown F: Perceived Difficulty (PD) Eliminating PP	Stepdown F: Perceived Performance (PP) Eliminating PD and Conf
Order	$F_{6,88} = .95$	.46	1.26	0.32	1.32
Knowledge	$F_{3,44} = 1.40$	.27	0.26	0.17	3.77
Order x Knowledge	$F_{6,88} = 0.74$	.62	0.89	0.65	0.72

Table 5

MANCOVA Summary

Number Right and Elimination Scores

with Facilitating, Debilitating, TRAIT and STATE as Covariates

	Multivariate		Stepdown F	
Source	F	P	NR	EL Eliminating NR
Order	$F_{4,172} = 0.33$	.86	0.28	0.38
Knowledge	$F_{2,86} = 0.02$	.98	0.01	0.02
Order x Knowledge	$F_{4,172} = 1.85$	.12	1.11	2.61

Table 6

PRE-POST STATE ANXIETY

Summary Table

Source	df	MS	F
Between			
Order	2	42.62	1.10
Knowledge	1	0.02	0.0005
Order x Knowledge	2	23.84	0.62
Error b	91	38.64	
Within			
Time	1	148.97	8.57**
Order x Time	2	42.11	2.42*
Knowledge x Time	1	2.01	0.12
Order x Knowledge x Time	2	25.40	1.46
Error w	91	17.38	

* $p < .10$

** $p < .05$

Table 7

Comparison of KR_{20} Reliabilities

for Experimental Forms, W^{+} values

(Reliabilities on Diagonal)

	1	2	3	4	5	6
1	.90					
2	1.25	.92				
3	1.20	1.50*	.88			
4	1.25	1.00	1.50*	.92		
5	1.25	1.00	1.50*	1.00	.92	
6	1.10	1.39*	1.09	1.39*	1.39*	.89

key 1 = Easy to Hard, Informed

2 = Uniform, Informed

3 = Random, Informed

4 = Easy to Hard, Not Informed

5 = Uniform, Not Informed

6 = Random, Not Informed

* $p < .05$

† $W = (1 - KR_{20(1)})/(1 - KR_{20(2)})$ is approximately distributed as F with df = $(N_1 - 1), (N_2 - 1)$

***Appendix

Test Instruments

[1]Spielberger's Self-Evaluation Questionnaire

STAI FORM X - 1 STATE

STAI FORM X - 2 TRAIT

[2]Test Experience Scale

(Alpert and Haber's Achievement Anxiety Scale)

Debilitating

Facilitating

[3] Mathematics Achievement Test Form II (Uniform)

Number Right

Elimination

Confidence

Perceived Difficulty

Perceived Performance

*** The Appendix has been removed because it contains copyrighted test instruments.

1. Spielberger, C. D.; Gorsuch, R. L.; Lushene, R. E. Manual for the State-Trait Anxiety Inventory, Palo Alto: Consulting Psychologist Press, 1970.

2. Alpert, R.; Haber, R. N. Anxiety in academic achievement situation, Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

3 Derived from the College Mathematics Placement Program, American College Testing Programs, Iowa City, Iowa, 1976.